

RESEARCH REPORT

ROBUST ESTIMATION OF THE VECTOR AUTOREGRESSIVE MODEL  
BY A TRIMMED LEAST SQUARES PROCEDURE

KRISTEL JOOSSENS • CHRISTOPHE CROUX

OR 0467

# Robust estimation of the vector autoregressive model by a trimmed least squares procedure

Kristel Joossens\*    and    Christophe Croux\*

K. U. Leuven

## Abstract

The vector autoregressive model is very popular for modeling multiple time series. Estimation of its parameters is done by a least squares procedure. However, this estimation method is unreliable when outliers are present in the data, and there is a need for robust alternatives. In this paper we propose to estimate the vector autoregressive model by using a trimmed least squares estimator. We show how the order of the autoregressive model can be determined in a robust way, and how confidence bounds around the robustly estimated impulse response functions can be constructed. The resistance of the estimators to outliers is studied on real and simulated data.

## 1 Introduction

The use of autoregressive models for predicting and modeling univariate time series is standard and well known. In many applications, one does not observe a single time series, but several series, possibly interacting with each other. For these multiple time series the vector autoregressive model has become very popular, and is described in standard textbooks on time series (e.g. Brockwell and Davis 2003, Chapter 7; Stock and Watson

---

\*Department of Applied Economics, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven, Belgium, E-mail: {kristel.joossens,christophe.croux}@econ.kuleuven.be.

2003, Chapter 14). In this paper we propose a robust procedure to estimate vector autoregressive models, to select their order, and to construct confidence bounds around the impulse response functions.

Let  $\{y_t \mid t \in \mathbb{Z}\}$  be a  $p$ -dimensional stationary time series. The vector autoregressive model of order  $k$ , denoted by  $\text{VAR}(k)$ , is given by

$$y_t = \mathcal{B}'_0 + \mathcal{B}'_1 y_{t-1} + \dots + \mathcal{B}'_k y_{t-k} + \varepsilon_t, \quad (1.1)$$

with  $y_t$  a  $p$ -dimensional vector, the intercept parameter  $\mathcal{B}'_0$  a vector in  $\mathbb{R}^p$  and the slope parameters  $\mathcal{B}_1, \dots, \mathcal{B}_k$  being matrices in  $\mathbb{R}^{p \times p}$ . Throughout the paper  $M'$  will stand for the transpose of the matrix  $M$ . The  $p$ -dimensional error terms  $\varepsilon_t$  are supposed to be independently and identically distributed with center zero and scatter matrix  $\Sigma$ . The latter means that the density of  $\varepsilon_t$  is of the form

$$f_{\varepsilon_t}(u) = \frac{g(u' \Sigma^{-1} u)}{(\det \Sigma)^{1/2}}, \quad (1.2)$$

with  $\Sigma$  a positive definite matrix, called the scatter matrix and  $g$  a positive function. If the second moment of  $\varepsilon_t$  exists, then  $\Sigma$  will be (proportional to) the covariance matrix of the error terms. Existence of a second moment, however, will not be required for the robust estimator. We focus on the unrestricted  $\text{VAR}(k)$  model, where no restrictions are put on the parameters  $\mathcal{B}_0, \mathcal{B}_1, \dots, \mathcal{B}_k$ .

Suppose that the multivariate time series  $y_t$  is observed for  $t = 1, \dots, T$ . The vector autoregressive model (1.1) can be rewritten as a multivariate regression model

$$y_t = \mathcal{B}' x_t + \varepsilon_t, \quad (1.3)$$

for  $t = k+1, \dots, T$  and with  $x_t = (1, y'_{t-1}, \dots, y'_{t-k})' \in \mathbb{R}^q$ , where  $q = pk + 1$ . The matrix  $\mathcal{B} = (\mathcal{B}'_0, \mathcal{B}'_1, \dots, \mathcal{B}'_k)' \in \mathbb{R}^{q \times p}$  contains all unknown regression coefficients. In the language of regression,  $X = (x_{k+1}, \dots, x_T)' \in \mathbb{R}^{n \times q}$  is the design matrix and  $Y = (y_{k+1}, \dots, y_T)' \in \mathbb{R}^{n \times p}$  the matrix of responses, where  $n = T - k$ . The classical least squares estimator for

the regression parameter  $\mathcal{B}$  in (1.3) is given by the well known formula

$$\hat{\mathcal{B}}_{\text{OLS}} = (X'X)^{-1}X'Y,$$

while the scatter matrix  $\Sigma$  is unbiasedly estimated by

$$\hat{\Sigma}_{\text{OLS}} = \frac{1}{n-p}(Y - X\hat{\mathcal{B}}_{\text{OLS}})'(Y - X\hat{\mathcal{B}}_{\text{OLS}}). \quad (1.4)$$

In applied time series research, one is aware of the fact that outliers can seriously affect parameter estimates, model specification and forecasts based on the selected model. For example, in business and economic data, it is common to see unusually low observations followed by unusually high observations or vice versa. Outliers in time series can be of different natures (Fox 1972), the most well known types being additive outliers and innovational outliers. If we consider the autoregressive model (1.1), then we say that  $y_t$  is an additive outlier if only its own value has been affected by contamination. On the other hand, we have an innovational outlier if the error term  $\varepsilon_t$  in (1.1) is contaminated. Innovational outliers will therefore have an effect on the next observations as well, due to the dynamic structure in the series. Additive outliers only have an isolated effect.

Several procedures to detect different types of outliers for univariate time series have been proposed (e.g. Chang, Tiao and Chen 1988, Gerlach, Carter and Kohn 1999 and Wu, Ravishanker and Hosking 1993). Bianco, García Ben, Martínez and Yohai (2001) and Riani (2004) propose diagnostic procedures based on robust estimators. Other robust estimators for univariate ARMA models have been proposed by Bustos and Yohai (1986) and De Luna and Genton (2001). All of the above studies focus on a single series.

The aim of this paper is to propose a robust estimation procedure for the multivariate autoregressive model, the most popular models for multiple time series analysis. Previous work on robust multivariate time series analysis uses estimators of the generalized M-type, which are known to have a low robustness in higher dimensions. Franses, Kloek and Lucas (1999) used generalized M-estimators and apply them on weekly scanning data. García Ben, Villar and Yohai (1999) use so-called residual autocovariance estimators, being an affine equivariant version of the estimators of Li and Hui (1989). Also the estimators of

García Ben et al (1999) are solutions of estimating equations. The latter may well have multiple solutions, which may not all correspond to robust solutions.

A common practice for handling outliers in a multivariate process is to first apply univariate techniques to the component series in order to remove the outlier, followed by treating the adjusted series as outlier-free and model them jointly. But this procedure encounters several difficulties. First, in a multivariate process, contamination in one component may be caused by an outlier in the other components. Secondly, a multivariate outlier cannot always be detected by looking at the component series separately, since it can be an outlier for the correlation structure only. Therefore it is better to cope with outliers in a multivariate framework. Tsay, Peña and Pankratz (2000) discuss the problem of multivariate outliers in detail.

To obtain a resistant estimator for the vector autoregressive model, we will replace the classical multivariate least squares estimator in (1.3) by a highly robust estimator. We will use the Multivariate Least Trimmed Squares estimator, introduced by Agulló, Croux and Van Aelst (2002). This estimator is defined by minimizing a trimmed sum of a squared Mahalanobis distances, and can be computed by a fast algorithm. The procedure also provides a natural estimator of the scatter of the residuals, which can then be used for model selection criteria. We present this estimator in Section 2. Section 3 explains how to determine in a robust way the autoregressive order of the model. In Section 4 confidence bounds around the impulse response functions obtained from the robust estimates are constructed. The robustness of the estimates is studied by means of several simulation experiments in Section 5, showing the effect of outliers on the estimation of the model and on the impulse response functions. In Section 6, the robust VAR methodology is applied on a real data set. Finally, Section 7 contains some conclusions.

## 2 The multivariate least trimmed squares estimator

The unknown parameters of the  $\text{VAR}(k)$  will be estimated via the multivariate regression model (1.3). For this the Multivariate Least Trimmed Squares estimator (MLTS), based on the idea of the Minimum Covariance Determinant estimator (Rousseeuw and Van

Driessen 1999), will be used. The MLTS selects the subset of  $h$  observations having the property that the determinant of the covariance matrix of its residuals from a least squares fit, solely based on this subset, is minimal.

Consider the data set  $Z = \{(x_t, y_t), t = k + 1, \dots, T\} \subset \mathbb{R}^{p+q}$ , and for any  $\mathcal{B} \in \mathbb{R}^{p \times q}$  denote  $r_t(\mathcal{B}) = y_t - \mathcal{B}'x_t$  the corresponding residuals. Let  $\mathcal{H} = \{H \subset \{k + 1, \dots, T\} \mid \#H = h\}$  be the collection of all subsets of size  $h$ . For any subset  $H \in \mathcal{H}$ , let  $\hat{\mathcal{B}}_{\text{OLS}}(H)$  be the classical least squares fit based on the observations of the subset

$$\hat{\mathcal{B}}_{\text{OLS}}(H) = (X_H' X_H)^{-1} X_H' Y_H,$$

where  $X_H$  and  $Y_H$  are submatrices of  $X$  and  $Y$ , consisting of the rows of  $X$ , respectively  $Y$ , having an index in  $H$ . The corresponding scatter matrix estimator computed from this subset is then

$$\hat{\Sigma}_{\text{OLS}}(H) = \frac{1}{h - p} (Y_H - X_H \hat{\mathcal{B}}_{\text{OLS}}(H))' (Y_H - X_H \hat{\mathcal{B}}_{\text{OLS}}(H)).$$

The MLTS estimator is now defined as

$$\hat{\mathcal{B}}_{\text{MLTS}}(Z) = \hat{\mathcal{B}}_{\text{OLS}}(\hat{H}) \quad \text{where} \quad \hat{H} = \underset{H \in \mathcal{H}}{\operatorname{argmin}} \det \hat{\Sigma}_{\text{OLS}}(H), \quad (2.1)$$

and the associated estimator of the scatter matrix of the error terms is given by

$$\hat{\Sigma}_{\text{MLTS}}(H) = \hat{\Sigma}_{\text{OLS}}(\hat{H}). \quad (2.2)$$

Equivalent characterizations of the MLTS estimator were given by Agulló, Croux and Van Aelst (2002). They proved that any  $\tilde{\mathcal{B}} \in \mathbb{R}^{p \times q}$  minimizing the sum of the  $h$  smallest squared Mahalanobis distances of its residuals (subject to  $\det \Sigma = 1$ ) is a solution of (2.1). In mathematical terms,

$$\hat{\mathcal{B}}_{\text{MLTS}} = \underset{\mathcal{B}, \Sigma; |\Sigma|=1}{\operatorname{argmin}} \sum_{s=1}^h d_{s:n}^2(\mathcal{B}, \Sigma).$$

Here  $d_{1:n}(\mathcal{B}, \Sigma) \leq \dots \leq d_{n:n}(\mathcal{B}, \Sigma)$  is the ordered sequence of the Mahalanobis distances  $d_s^2(\mathcal{B}, \Sigma) = r_s(\mathcal{B})' \Sigma^{-1} r_s(\mathcal{B})$ . Therefore, we see that the MLTS-estimator minimizes the sum of the  $h$  smallest squared distances of its residuals, and is therefore the multivariate extension of the Least Trimmed Squares (LTS) estimator of Rousseeuw (1984).

Since the efficiency of the MLTS estimator is rather low, we will use the reweighted version to improve the performance of the MLTS. If  $\hat{\mathcal{B}}_{\text{MLTS}}$  and  $\hat{\Sigma}_{\text{MLTS}}$  denote the initial MLTS estimates, then the one-step Reweighted Multivariate Least Trimmed Squares (RMLTS) estimates are defined as

$$\hat{\mathcal{B}}_{\text{RMLTS}} = \hat{\mathcal{B}}_{\text{OLS}}(J) \quad \text{and} \quad \hat{\Sigma}_{\text{RMLTS}} = c_\delta \hat{\Sigma}_{\text{OLS}}(J), \quad (2.3)$$

where  $J = \{j \mid d_j^2(\hat{\mathcal{B}}_{\text{MLTS}}, \hat{\Sigma}_{\text{MLTS}}) \leq q_\delta\}$  and  $q_\delta = \chi_{q,1-\delta}^2$  is the upper  $\delta$ -quantile of a  $\chi^2$  distribution with  $q$  degrees of freedom. Here  $\delta$  is the trimming fraction in the reweighting step and  $c_\delta$  is a consistency factor to obtain consistent estimation of  $\Sigma$  at the model distribution (1.2) of the error terms. In the case of multivariate normal error terms it has been shown (e.g. Croux and Haesbroeck 1999) that  $c_\delta = (1 - \delta)/F_{\chi_{p+2}^2}(q_\delta)$ . Here  $F_{\chi_q^2}$  is the cumulative distribution function of a  $\chi^2$  distribution with  $q$  degrees of freedom. The idea is that outliers have large residuals with respect to the initial robust MLTS estimator, resulting in a large residual Mahalanobis distance  $d_j^2(\hat{\mathcal{B}}_{\text{MLTS}}, \hat{\Sigma}_{\text{MLTS}})$ . If this Mahalanobis distance is above the critical value  $\chi_{q,1-\delta}^2$ , then the observation is flagged as an outlier. The final RMLTS is based on the observations not detected as outliers. In this paper, we set  $\delta = 0.01$  and take as trimming portion for the initial MLTS estimator  $1 - h/n \approx 25\% = \alpha$ .

### 3 Determining the Autoregressive Order

To select the order of the vector autoregressive model, criteria for different orders are computed and an estimator of the optimal order is obtained by minimizing this criterion. Most of these criteria are in terms of the value of the log likelihood  $l_k$  of the VAR( $k$ ).

Using the model hypothesis on the error terms, we get

$$l_k = \sum_{t=k+1}^T g(\varepsilon_t' \Sigma^{-1} \varepsilon_t) - \frac{n}{2} \log \det \Sigma.$$

Recall that  $n = T - k$ . When error terms are multivariate normal the above leads to

$$l_k = -\frac{n}{2} \log \det \Sigma - \frac{np}{2} \log(2\pi) - \frac{1}{2} \sum_{t=k+1}^T \varepsilon_t' \Sigma^{-1} \varepsilon_t. \quad (3.1)$$

The log likelihood will depend on the autoregressive order via the estimate of the covariance matrix of the residuals. For the ordinary least squares estimator we have

$$\hat{\Sigma}_{\text{OLS}} = \frac{1}{n-p} \sum_{t=k+1}^T \hat{\varepsilon}_t(k) \hat{\varepsilon}_t'(k),$$

where the  $\hat{\varepsilon}_t(k)$  are the residuals corresponding with the estimated vector autoregressive model of order  $k$ . Using the trace properties, we have that the last term in (3.1) equals the constant  $-(n-p)p/2$  for the OLS estimator. To prevent that outliers might affect the estimation of the VAR models and therefore also the estimation of the criteria, we will estimate  $\Sigma$  by the robust reweighted multivariate least trimmed squares estimator instead of the ordinary least squares estimator. When using the RMLTS, discussed in Section 2, we get

$$\hat{\Sigma}_{\text{RMLTS}} = \frac{c_\delta}{m-p} \sum_{t \in J} \hat{\varepsilon}_t(k) \hat{\varepsilon}_t'(k),$$

with  $J$  as in (2.3) and  $m$  the number of elements in  $J$ . Now, the last term in (3.1) becomes  $-(m-p)p/2c_\delta$ , where the value of  $m$  might depend on the selected order  $k$ .

We will consider the following three criteria: the popular Akaike info criterion (Akaike 1973) defined as

$$\text{AIC}(k) = \frac{-2}{n} l_k + \frac{2}{n} (kp + 1)p,$$

the Hannan-Quinn criterion (Hannan & Quinn 1979) given by

$$\text{HQ}(k) = \frac{-2}{n} l_k + \frac{2 \log(\log(n))}{n} (kp + 1)p,$$



and the Schwarz criterion (Schwarz 1978) which equals

$$SC(k) = \frac{-2}{n}l_k + \frac{\log(n)}{n}(kp + 1)p.$$

In all three criteria,  $(kp + 1)p$  is the number of unknown regression parameters and penalizes for model complexity.

To illustrate that outliers affect the lag length criteria, we use the bivariate time series “gsln3” (Tsay 2002, p. 324–325). The first series “gsl” is the 1-year Treasury constant maturity rate, and the second series “n3” is the 3-year Treasury constant maturity rate. The data are monthly and obtained from the Federal Reserve Bank of St Louis. The sampling period is from April 1953 to January 2001. As in the book of Tsay (2002), we work with the log-transformed versions of both series. From the plot of the two log transformed series (Figure 1), it can be seen that there might be some outliers around the years 1954 and 1958.

In Table 1 the lag length criteria using the ordinary least squares estimator and the reweighted multivariate trimmed least squares estimator are represented. The information criteria clearly depend on the chosen estimator. For example, when using the AIC the classical method suggests a VAR(8) model while the robust indicates a VAR(6) model. On the other hand the Schwarz criterion selects an optimal order 3 for both estimators. Since the latter criterion yields a consistent estimate of the optimal order (Hannan 1980) we continue the analysis with  $k = 3$ .

After estimating the VAR(3) model with the robust RMLTS estimator, the corresponding robust residual distances  $d_t(\hat{\mathcal{B}}_{\text{RMLTS}}, \hat{\Sigma}_{\text{RMLTS}})$  are computed. Figure 2 displays these distances with respect to the time index, and indicates outlying observations. Since residual squared distances follow a  $\chi_p^2$  distribution, it is common to compare these distances with the critical value  $\chi_{p,0.99}$  (Rousseeuw and Van Zomeren 1991). Figure 2 reveals that several suspectable high residuals are detected, indicating presence of huge outliers around the years 1954 and 1958. Therefore it seems appropriate to make use of robust methods for further analysis of this data set.

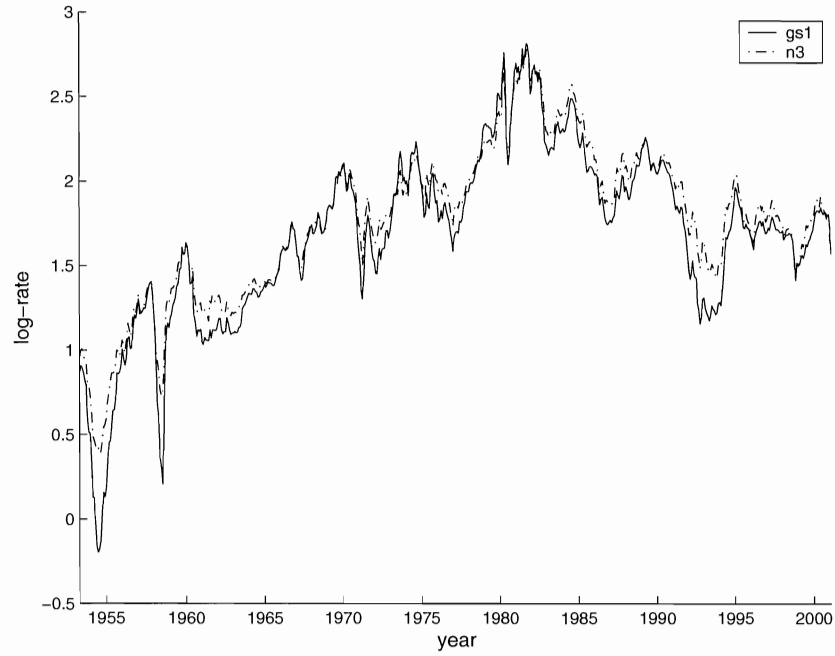


Figure 1: Time plot of the log U.S. monthly interest rates from April 1953 to January 2001. The solid line represents the 1-Year Treasury constant maturity rate and the dot-dashed line represents the 3-Year Treasury constant maturity rate.

Table 1: Lag length criteria using the OLS and RMLTS estimator for the “gs1n3” series.

$k$	1	2	3	4	5	6	7	8
Based on OLS estimation								
AIC	-7.3552	-7.5823	-7.6179	-7.6261	-7.6149	-7.6078	-7.6268	<b>-7.6276</b>
HQ	-7.3374	-7.5526	<b>-7.5763</b>	-7.5726	-7.5494	-7.5302	-7.5372	-7.5258
SC	-7.3096	-7.5062	<b>-7.5113</b>	-7.4889	-7.4470	-7.4090	-7.3972	-7.3669
Based on RMLTS estimation								
AIC	-7.4386	-7.6282	-7.6795	-7.6997	-7.6943	<b>-7.7490</b>	-7.6961	-7.7118
HQ	-7.4208	-7.5985	-7.6380	-7.6461	-7.6288	<b>-7.6714</b>	-7.6065	-7.6101
SC	-7.3930	-7.5522	<b>-7.5730</b>	-7.5624	-7.5264	-7.5502	-7.4665	-7.4512

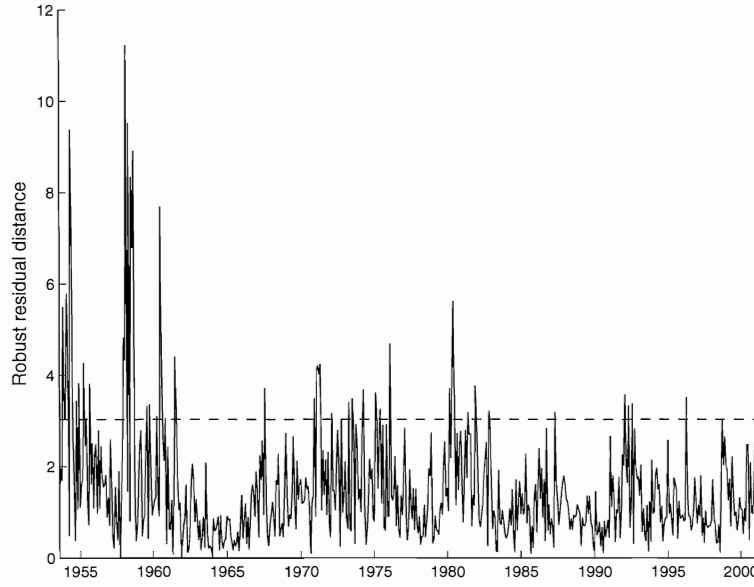


Figure 2: Robust residual distances for the “gsln3” series based on the robust RMLTS estimator of a VAR(3) model. The dashed line represents the critical value  $\chi_{2,0.99} = 3.0349$ .

## 4 Impulse response function

After selecting and estimating the VAR model, it is common and instructive to look at the Impulse response Function (IRF), which permits to quantify variable responses to shocks on different horizon lengths (Hamilton 1994, chapter 11). The impulse response functions are defined as follows. Let  $\mathcal{B}(L)$  be the autoregressive polynomial  $I_p - \mathcal{B}'_1 L \dots - \mathcal{B}'_k L^k$ . We can write the vector autoregressive model as  $\mathcal{B}(L)y_t = \mathcal{B}'_0 + \varepsilon_t$ , or  $y_t = \mathcal{B}(L)^{-1}\mathcal{B}'_0 + \mathcal{B}(L)^{-1}\varepsilon_t$ , which can be rewritten as an infinite moving average model

$$y_t = a + \varepsilon_t + \mathcal{A}_1 \varepsilon_{t-1} + \dots + \mathcal{A}_l \varepsilon_{t-l} + \dots$$

where  $a$  is a vector and  $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_l, \dots$  are  $p \times p$  matrices, depending on the parameters in  $\mathcal{B}$ . The function mapping  $l$  on  $(\mathcal{A}_l)_{ij}$ , for  $l = 1, 2, 3, \dots$  is called an impulse response function. It measures the response of component  $i$  of  $y_t$  to an impulse of one unit in component  $j$  of  $\varepsilon_{t-l}$ . Letting the indices  $i$  and  $j$  range between 1 and  $p$ , results in  $p^2$  impulse response functions. In practice  $\mathcal{B}$  is estimated, yielding an estimated IRF. Using

a robust estimate of  $\mathcal{B}$ , yields then a robustly estimated IRF. In the sequel, we consider two methods to construct confidence bounds around these estimated IRFs.

To construct the analytic confidence bounds, we start from the asymptotic normality of the estimators in the multivariate regression model (1.3):

$$\sqrt{T}(\text{vec}\hat{\mathcal{B}} - \text{vec}\mathcal{B}) \rightarrow N_{p(pk+1)}(0, c_p \Sigma \otimes Q^{-1}). \quad (4.1)$$

Here  $Q = E[x_t x_t']$ , with  $x_t$  as in Section 1, and “vec” is the operator which vectorizes a matrix and  $\otimes$  stands for the kronecker product. The constant  $c_p$  depends on the chosen estimator. For the OLS estimator we have  $c_p = 1$ , and for the RMLTS the value of  $c_p$  will be larger than 1 and can be retrieved from the asymptotic variance of the RMLTS estimator for the multivariate regression model (Agulló, Croux and Van Aelst 2002). Note that  $c_p$  not only depends on the dimension, but also depends on the trimming fraction  $\alpha$  of the initial MLTS estimator and on the trimming fraction  $\delta$  used in the reweighting step. For the RMLTS we have

$$c_p = c_p(\alpha, \delta) = \frac{c^2}{F_\alpha} + \frac{F_\delta}{(1 - \delta)^2} + \frac{c}{1 - \delta} \frac{F_\delta}{F_\alpha}$$

with  $c = 1 - F_\delta/(1 - \delta)$  where  $F_\delta = F_{\chi_{p+2}^2}(\chi_{p,1-\delta}^2)$  and  $F_\alpha = F_{\chi_{p+2}^2}(\chi_{p,1-\alpha}^2)$ .

By using the Delta method, as in Hamilton (1994, page. 186), we get from (4.1) that

$$\sqrt{T}(\text{vec}\mathcal{A}_l(\hat{\mathcal{B}}) - \text{vec}\mathcal{A}_l(\mathcal{B})) \rightarrow N_{p^2}(0, c_p G_l(\Sigma \otimes Q^{-1}) G_l'),$$

where

$$G_l = \frac{\partial \text{vec}\mathcal{A}_l(\mathcal{B})}{\partial (\text{vec}\mathcal{B})'}.$$

Standard errors around the values  $(\mathcal{A}_l)_{ij}$  of the IRFs are then obtained as the square roots of the diagonal elements of  $c_p \frac{1}{T} \hat{G}_l(\hat{\Sigma} \otimes \hat{Q}^{-1}) \hat{G}_l'$ . Here we take  $\hat{\Sigma}$  as in (2.3), and  $\hat{Q} = \text{average}_{t \in J} x_t x_t'$  with  $J$  the set of indices used in the definition of the RMLTS estimator. The matrix  $G_l$  is estimated using the expression of Hamilton (1994, p. 337), who showed

that  $G_l$  can be calculated recursively via the formula

$$G_l = \sum_{s=1}^l [\mathcal{A}_{s-1} \otimes (O_{n1} \mathcal{A}_{l-s} \mathcal{A}_{l-s-1} \dots \mathcal{A}_{l-s-k-1})],$$

where  $O_{nm}$  is a zero matrix of size  $n \times m$ . The matrices  $\mathcal{A}_l$  in the above expressions need then to be estimated by  $\mathcal{A}_l(\hat{\mathcal{B}})$ .

We can also obtain Monte Carlo confidence bounds using a parametric bootstrap procedure. We first estimate the model from the original data. Then we generate 1000 series according to the estimated VAR( $k$ ) model, with errors following a multivariate normal distribution with mean zero and covariance matrix  $\hat{\Sigma}$ . From these 1000 generated series, 1000 impulse response values can be computed. By sorting these values and taking the 2.5% and 97.5% quantile, 95% confidence bounds can be constructed for the impulse response functions.

The analytic and Monte Carlo confidence intervals around the IRFs estimated by the robust procedure have been computed for the “gs1n3”-data containing the Maturity rates GS1 and N3 (see Figure 3) for the robust RMLTS estimator. We see that the impulse response functions for the analytic method and the Monte Carlo method are very similar. The Monte Carlo based 95% confidence bounds are marginally larger and are less smooth in comparison with the analytic confidence bounds.

It is seen from Figure 3 that the effect of a unit shock at the innovation of GS1 on the response of GS1 is significant up to 11 months. On the other hand the variable N3 is non-responsive to such a shock. The effects of a unit shock in the innovation series driving N3 are more important: the response of GS1 is significant to almost 3 years and the response of N3 is significant up to almost 4 years.

## 5 Robustness of the estimator

In order to look at the robustness of the estimates, we will perform a simulation study comparing the OLS estimator with the robust RMLTS estimator. Consider a bivariate

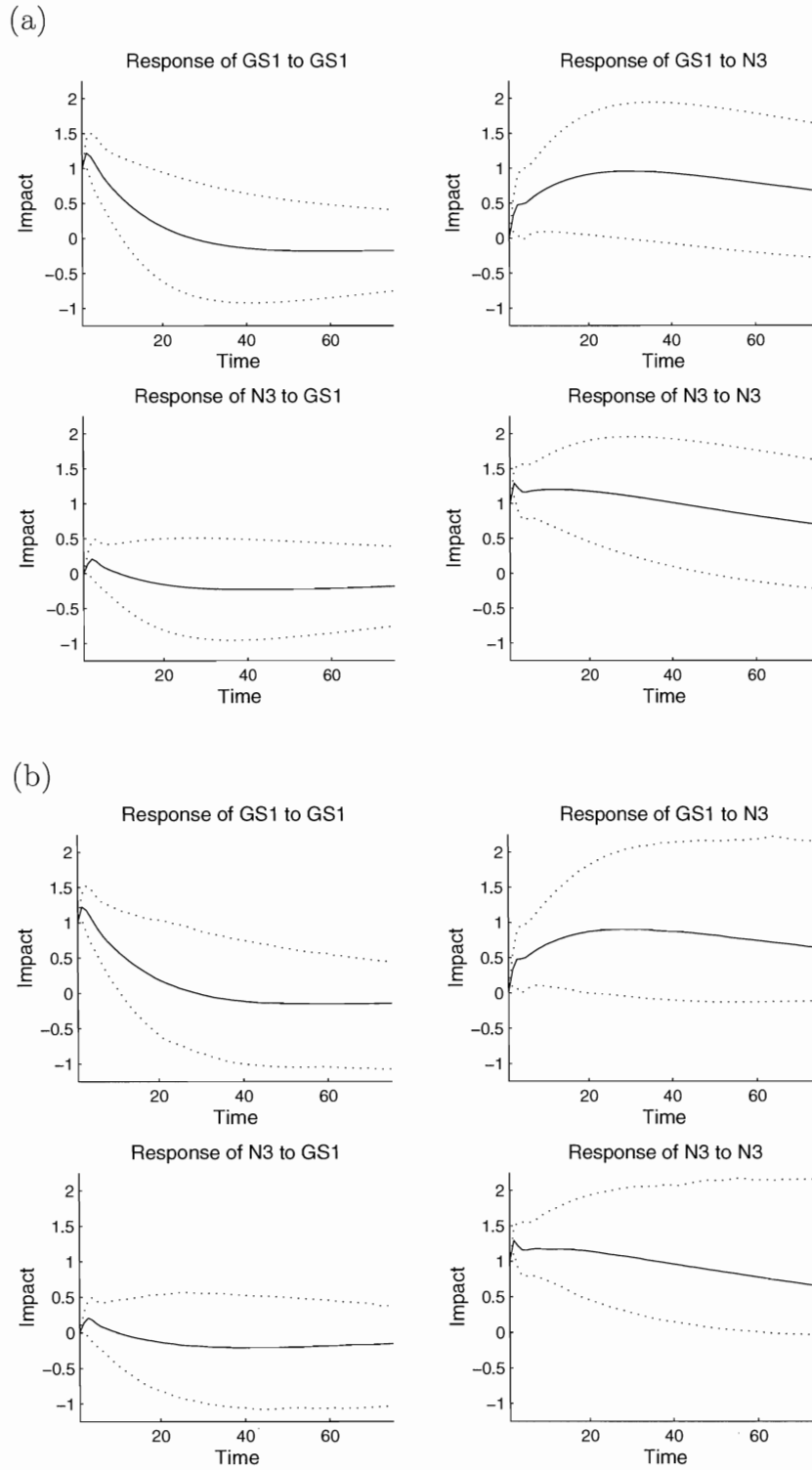


Figure 3: The impulse response functions (solid lines) for the “gs1n3” data based on VAR(3) models estimated by the robust RMLTS estimator, together with its (a) analytic; (b) Monte Carlo confidence bounds (dotted lines).

time series generated according the VAR(2) model

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} .10 \\ .02 \end{pmatrix} + \begin{pmatrix} .40 & .03 \\ .04 & .20 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} .100 & .005 \\ .010 & .080 \end{pmatrix} \begin{pmatrix} y_{1,t-2} \\ y_{2,t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}, \quad (5.1)$$

where  $\varepsilon_t \sim N_2(0, \Sigma)$  with

$$\Sigma = \begin{pmatrix} 1 & .2 \\ .2 & 1 \end{pmatrix}. \quad (5.2)$$

The aim is to look at the effect of the outliers on the parameter estimates. Since there are 10 parameters to be estimated for each equation, we look at the total bias and total Mean Squared Error (MSE) as performance measures. The bias is computed as

$$\text{Bias} = \sqrt{\sum_{i=1}^q \sum_{j=1}^p \left( \frac{1}{\text{nsim}} \sum_{s=1}^{\text{nsim}} \hat{\mathcal{B}}_{ij}^s - \mathcal{B}_{ij} \right)^2},$$

where  $\hat{\mathcal{B}}^s$ , for  $s = 1, \dots, \text{nsim}$ , is the estimate obtained from the  $s$ -th generated series,  $\mathcal{B}$  is the true parameter value and  $\text{nsim} = 1000$  the number of simulations. The MSE equals

$$\text{MSE} = \sum_{i=1}^q \sum_{j=1}^p \left[ \frac{1}{\text{nsim}} \sum_{s=1}^{\text{nsim}} (\hat{\mathcal{B}}_{ij}^s - \mathcal{B}_{ij})^2 \right].$$

The classical and robust estimators are used to estimate the VAR(2) model for the uncontaminated series and the contaminated series, where outliers are added.

## 5.1 Additive outliers

After generating the series of length  $T = 500$ , additive outliers are introduced. The contamination is done by randomly selecting  $m$  bivariate observations, and adding the value 10 to all the components of the selected observations. We have considered different contamination levels, ranging from one single up to 40 additive outliers.

The Bias and MSE for the classical OLS estimator and the robust RMLTS estimator are given in Table 2, as a function of the number  $m$  of additive outliers. We see that

Table 2: Simulated Bias and Mean Squared error for the classical (OLS) and robust (RMLTS) estimator of a bivariate VAR(2) model, in presence of  $m$  additive outliers in a series of length 500.

$m$	OLS		RMLTS	
	Bias	MSE	Bias	MSE
0	0.0076	0.0208	0.0080	0.0224
1	0.0868	0.0303	0.0202	0.0232
2	0.1453	0.0467	0.0311	0.0245
3	0.1879	0.0625	0.0408	0.0256
4	0.2221	0.0790	0.0484	0.0271
5	0.2546	0.0967	0.0543	0.0278
10	0.3905	0.1967	0.0758	0.0387
15	0.5162	0.3201	0.1129	0.0567
20	0.6390	0.4716	0.1762	0.0801
25	0.7701	0.6666	0.2376	0.1011
30	0.8924	0.8770	0.2895	0.1174
35	1.0154	1.1251	0.3137	0.1263
40	1.1454	1.4133	0.3271	0.1308

the Bias and MSE grows for an increasing number of outliers. The increase in Bias and MSE is much faster for the method using OLS. Using the robust estimator instead of OLS leads to a small loss in efficiency at the model when no outliers are present. When only even one outlier is introduced, the RMLTS is already more efficient, and the gain in MSE for the RMLS becomes highly substantial for the larger amounts of outliers. associated impulse response functions.

The robustness of the estimates of the impulse response function is studied as well. For each of the  $n_{sim}=1000$  simulated series, the IRF is computed. A plot of the averaged impulse response functions, together with the 5% and 95% quantiles over all the generated IRFs is made in Figure 4. This figure represents the impulse response functions of the response of Variable 1 to a unit shock in the innovation term of Variable 2, once as estimated by using ordinary least squares estimators, and once by using the reweighted multivariate least trimmed squares estimators. We hardly see any difference between the IRF based on the classical OLS estimator and the one based on robust RMLTS estimator,



but the quantiles are somehow wider while using the robust estimator, since the RMLTS is a bit less efficient when no outliers are present. We now add 5 additive outliers in order to look at their effect on the impulse response functions. From Figure 5 we see that the average IRFs using the robust estimator looks more like the average IRFs of the uncontaminated data in Figure 4. But, when using the classical estimator, there is a change of the shape in the impulse response function. Hence, and without much surprise, the non-robustness of the OLS estimator is also reflected in the

## 5.2 Innovational outliers

Instead of contaminating the series with additive outliers, we look now at the effect of innovational outliers. To generate these innovational outliers, we randomly select a number  $m$  of innovation terms in (5.1) and add 10 to the first component of the innovations, to come to the contaminated innovations series  $\varepsilon_t^C$ . These are then used to generate the bivariate series according to (5.1), but with  $\varepsilon_t$  replaced by  $\varepsilon_t^C$ . Recall that innovational outliers have a more persistent effect on the series  $y_t$  than additive outliers. The Bias and MSE when estimating the uncontaminated ( $m = 0$ ) and contaminated series are given in Table 3, for as well the classical as the robust estimation.

The Bias and MSE for OLS grow quite fast for an increasing number of innovational outliers, although at a smaller rate as for contamination with additive outliers. For the robust estimator we see a small decrease of the Bias and MSE, implying that the robust procedure becomes more efficient in presence of innovational outliers. This is due to the fact that an innovational outlier in the time series result in a single vertical outlier, but also in several good leverage points when estimating the autoregressive model. The robust method can cope with the vertical outlier and takes profit of the good leverage points to decrease the MSE. The OLS estimator gets biased due to the vertical outliers, but the presence of the good leverage points explains why the effect of innovational outliers is less strong as for additive outliers. Note that when no outliers are present, the RMLTS is again very close to OLS, the loss in efficiency being marginal.

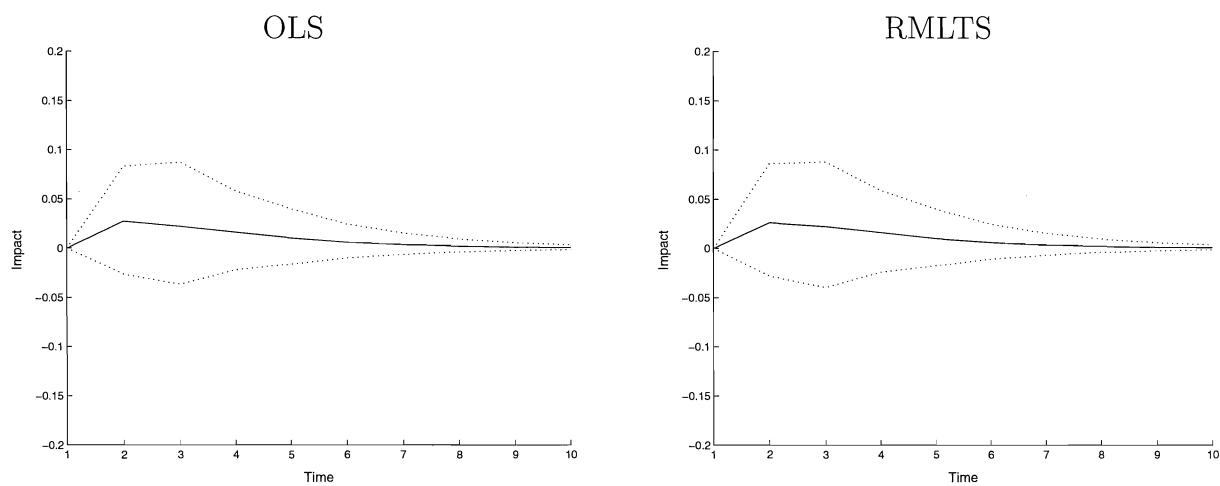


Figure 4: Simulated average impulse response functions (solid lines) and the 5% and 95% quantiles (dotted lines) for a bivariate VAR(2) model using the OLS estimator (left) and the RMLTS estimator (right).

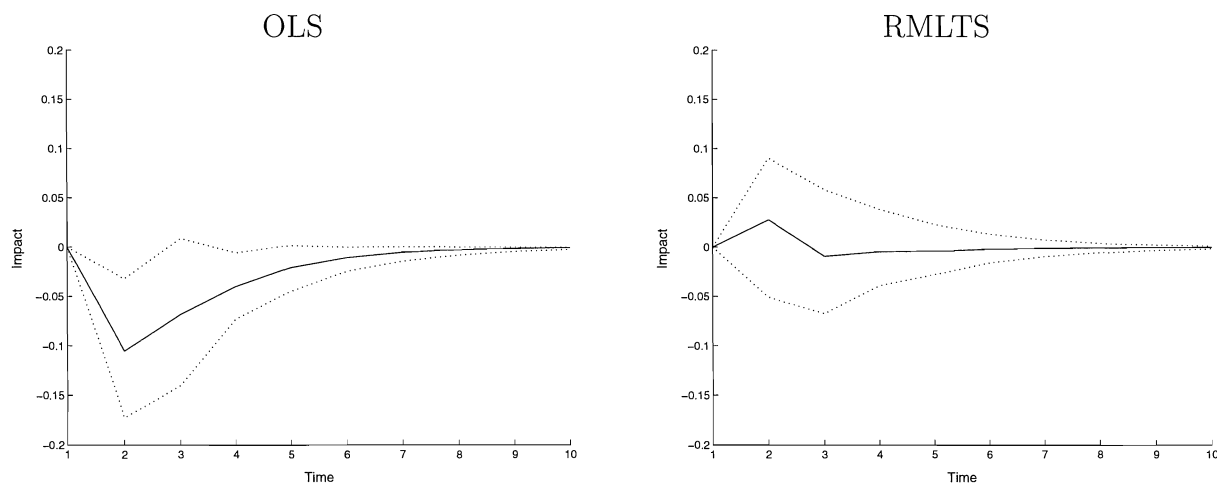


Figure 5: As in Figure 4, but now after adding 5 additive outliers (out of 500 observations).

Table 3: Bias and Mean Squared error for the classical (OLS) and robust (RMLTS) estimator of a bivariate VAR(2) model, in presence of  $m$  innovational outliers in a series of length 500.

$m$	OLS		RMLTS	
	Bias	MSE	Bias	MSE
0	0.0086	0.0216	0.0093	0.0234
1	0.0230	0.0218	0.0085	0.0219
2	0.0424	0.0232	0.0080	0.0210
3	0.0629	0.0264	0.0088	0.0202
4	0.0830	0.0298	0.0089	0.0197
5	0.1028	0.0335	0.0079	0.0190
10	0.2017	0.0685	0.0076	0.0176
15	0.3024	0.1226	0.0082	0.0172
20	0.3977	0.1935	0.0071	0.0168
25	0.4975	0.2875	0.0071	0.0169
30	0.5915	0.3939	0.0072	0.0171
35	0.6878	0.5228	0.0072	0.0169
40	0.7783	0.6622	0.0072	0.0171

## 6 Example

We illustrate the use of the robust estimation of the parameters of a vector autoregressive model on a real time series, called “housing” data (Diebold 2001, p. 109). Housing is a bivariate series of monthly data of housing starts and housing completions from January 1968 up to June 1996. To find the optimal order of the vector autoregressive model, the robust lag length criteria are used and each of them suggests a VAR(4) model. After estimating the model, robust residuals distances are computed and plotted in Figure 6. Two severe outliers are detected, together with two other extreme observations. Due to the presence of these atypical observations, it is indeed advisable to use a robust estimation procedure.

Impulse response functions resulting from the estimated VAR(4) model using OLS and RMLTS are presented in Figure 7. In all four IRF we see that the effect of a shock is underestimated when the OLS estimator is used with respect to using the RMLTS estimator. The robust IRFs are also somewhat smoother. A unit shock in the innovation

for the variable "housing starts" gives rise to a large response on the completions series (and also on its housing starts series). On the other hand, a unit shock in the innovations of the completions series has only a limited impact on both series. Since outliers are present in the data, we think that more confidence should be given to the robust analysis.

## 7 Conclusion

For multivariate time series it may occur that correlation outliers can be present, which are not necessarily visible in plots of the single univariate series. Development of robust procedures for multiple time series analysis is therefore even more important than for univariate time series analysis.

In this paper we show how robust multivariate regression estimators can be used to estimate Vector Autoregressive models. Focus was on the reweighted multivariate least squares estimators, since the objective function defining this estimator is closely related to several information criteria. Of course, other robust multivariate regression estimators can be used as well (e.g. the MM estimators of Tatsuoka and Tyler 2000 or the robust covariance based estimators of Rousseeuw et al 2004). Software to robustly estimate the VAR model is available at <http://www.econ.kuleuven.be/kristel.joossens>, and was used to analyze the real data set of Section 6. This software provides robust distances as a tool for outlier detection, computes different robust lag-length criteria, as well as robustly estimated impulse response functions. These impulse response functions are used in applied multivariate time series especially in economics for interpreting the estimated VAR model. We also provide the confidence bounds around this robust impulse response functions..

The simulation experiments in Section 5 clearly illustrated the advantage of using the reweighted multivariate least trimmed squares estimator instead of the classical approach. If there are no outliers in the data set present, the robust estimator performs almost as good as the classical estimator. But if there are outliers, then bias and MSE remain under control when using the robust estimator.

**Acknowledgment:** We wish to thank Prof. M. Dekimpe for useful suggestions. This

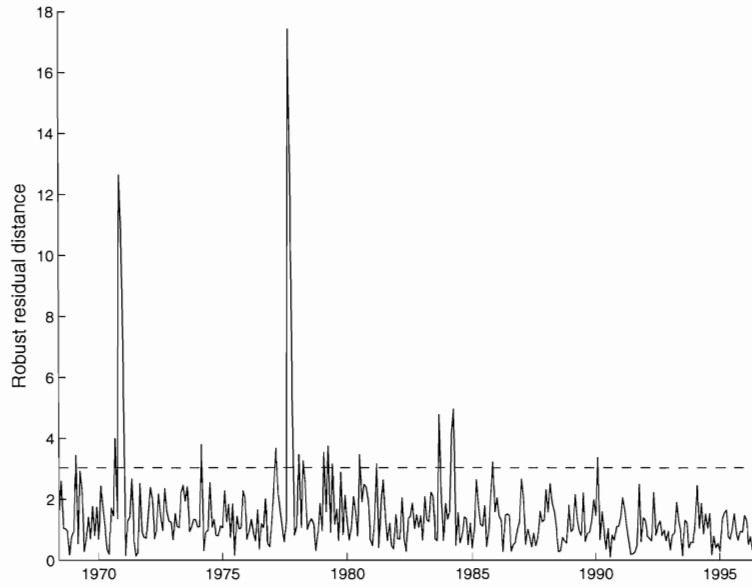


Figure 6: Robust residual distances for the housing data set. The dashed line represents the critical value.

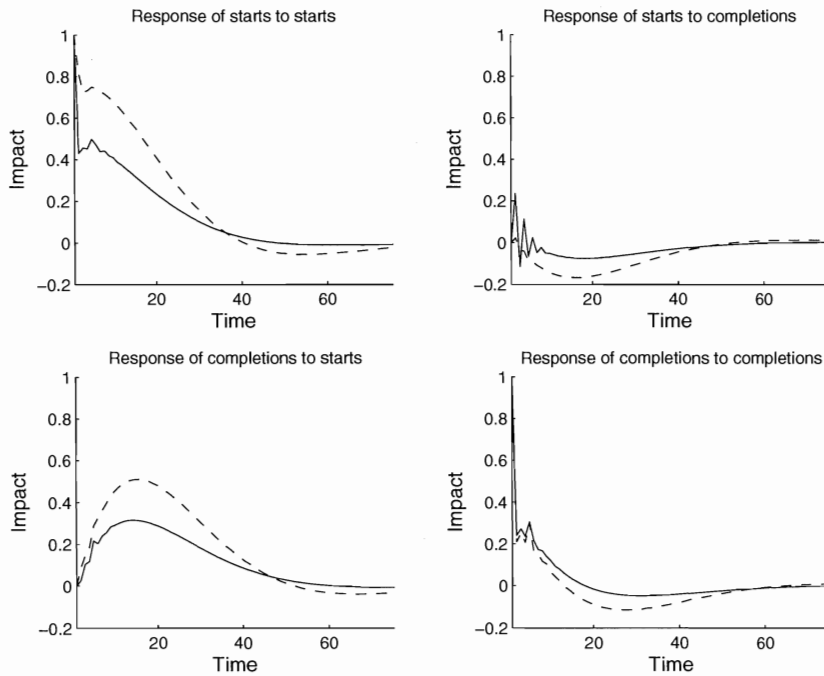


Figure 7: The impulse response functions for the housing data set, based on a VAR(4) model estimated by the classical OLS estimator (solid line) and the robust RMLTS estimator. The variables in the housing data set are starts and completions.

research has been supported by the Research Fund K.U. Leuven and the “Fonds voor Wetenschappelijk Onderzoek” (Contract number G.0385.03).

## References

- Agulló, J., Croux, C., and Van Aelst, S. (2002), “The multivariate least trimmed squares estimator,” *Research report*, Dept. of Applied Economics, K.U.Leuven, Belgium.
- Akaike, H. (1973), “Information theory and an extension of the maximum likelihood principle,” in *2nd International Symposium on Information Theory*, eds. Petrov, B. N. and Csáki, F., Akadémiai Kiadó: Budapest, pp. 267–281.
- Bianco, A. M., García Ben, M., Martínez, E. J., and Yohai, V. J. (2001), “Outlier detection in regression models with ARIMA errors using robust estimates,” *Journal of Forecasting*, Vol. 20, 565–579.
- Brockwell, P. J. and Davis, R. A. (2003), *Introduction to Time Series and Forecasting*, Wiley: New York, 2nd ed.
- Bustos, H. and Yohai, V. J. (1986), “Robust estimates of ARMA models,” *Journal of the American Statistical Association*, Vol. 81(393), 155–159.
- Chang, I., Tiao, G. C., and Chen, C. (1988), “Estimation of time series parameters in the presence of outliers,” *Technometrics*, Vol. 30, 193–204.
- Croux, C. and Haesbroeck, G. (1999), “Influence function and efficiency of the MCD-scatter matrix estimator,” *Journal of Multivariate Analysis*, Vol. 71, 161–190.
- De Luna, X. and Genton, M. G. (1967), “Robust simulation-based estimation of ARMA models,” *Journal of Computational and Graphical Statistics*, Vol. 10(2), 370–387.
- Diebold, F. X. (2001), *Elements of Forecasting*, South Western College Publishing, 2nd ed.

- Fox, A. (1972), "Outliers in time series," *Journal of the Royal Statistical Society, Series B*, Vol. 34(3), 350–363.
- Franses, H., Kloek, T., and Lucas, A. (1999), "Outlier robust analysis of long-run marketing effects for weekly scanning data," *Journal of Econometrics*, Vol. 89, 293–315.
- García Ben, M., Martínez, E. J., and Yohai, V. J. (1999), "Robust estimation in vector autoregressive moving average models," *Journal of Time Series Analysis*, Vol. 20(4), 381–399.
- Gerlach, R., Carter, C., and Kohn, R. (1999), "Diagnostics for time series analysis," *Journal of Time Series Analysis*, Vol. 20(3), 309–330.
- Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.
- Hannan, E. J. (1980), "The estimation of the order of an ARMA process," *Annals of Statistics*, Vol. 8, 1071–1081.
- Hannan, E. J. and Quinn, B. G. (1979), "The determination of the order of an autoregression," *Journal of the Royal Statistical Society, Series B*, Vol. 41(2), 190–195.
- Li, W. K. and Hui, Y. V. (1989), "Robust multiple Time series modelling," *Biometrika*, Vol. 76(2), 309–315.
- Riani, M. (2004), "Extensions of the forward search to time series," *Studies in Non Linear Dynamics and Econometrics*, Vol. 8(2), Article 2.
- Rousseeuw, P. J. (1984), "Least median of squares regression," *Journal of the American Statistical Association*, Vol. 79, 871–880.
- Rousseeuw, P. J. and Van Driessen, K. (1999), "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, Vol. 41, 212–223.
- Rousseeuw, P. J., Van Driessen, K., Van Aelst, S., and Agulló, J. (2004), "Robust multivariate regression," *Technometrics*, Vol. 46(3), 293–305.

- Rousseeuw, P. J. and Van Zomeren, B. C. (1991), “Robust distances: simulations and cutoff values,” in *Directions in Robust Statistics and Diagnostics, part II*, eds. Stahel, W. and Weisberg, S., Springer Verlag: New York, vol. 34 of *the IMA Volumes in Mathematics and Its Applications*, pp. 195–203.
- Schwarz, G. (1978), “Estimating the dimension of a model,” *Annals of Statistics*, Vol. 6(2), 461–464.
- Stock, J. H. and Watson, M. W. (2003), *Introduction to Econometrics*, Addison Wesley.
- Tatsuoka, K. S. and Tyler, D. E. (2000), “On the uniqueness of the  $S$ -functionals and the  $M$ -functionals under nonelliptical distributions,” *The Annals of Statistics*, Vol. 28, 1219–1243.
- Tsay, R. S. (2002), *Analysis of Financial Time Series*, John Wiley & Sons: New York.
- Tsay, R. S., Peña, D., and Pankratz, A. E. (2000), “Outliers in multivariate time series,” *Biometrika*, Vol. 87(4), 789–804.
- Wu, L. S.-Y., Ravishanker, N., and Hosking, J. R. M. (1993), “Reallocation outliers in time series,” *Applied Statistics*, Vol. 42(2), 301–313.



